Transporter's evolution and carbohydrate metabolic clusters

Plantinga, Titia H.; Does, Chris van der; Driessen, Arnold J.M.

Published in:
Trends in Microbiology

DOI:
10.1016/j.tim.2003.11.001

Publication date:
2004

16 Collins, R.F. *et al.* (2003) *Three-dimensional structure of the* Neisseria meningitidis *secretin PilQ determined from negative stain transmission electron microscopy. J. Bacteriol.* 185, 2611–2617
17 Spratt, B.G. *et al.* (2001) The relative contributions of recombination and point mutation to the diversification of bacterial clones. *Curr. Opin. Microbiol.* 4, 602–606
18 Bjedov, I. *et al.* (2003) Stress-induced mutagenesis in bacteria. *Science* 300, 1404–1409
19 Rad, M.E. *et al.* (2003) Mutations in putative mutator genes of *Mycobacterium tuberculosis* strains of the W-Beijing family. *Emerg. Infect. Dis.* 9, 838–845
20 Rainey, B.P. and Moxon, E.R. (2000) When being hyper keeps you fit. *Science* 288, 1186–1187
21 Bayliss, C.D. *et al.* (2002) Mutations in *polI* but not *mutSLH* destabilize *Haemophilus influenzae* tetranucleotide repeats. *EMBO J.* 21, 1465–1476
22 Martin, P. *et al.* (2003) Experimentally revised repertoire of putative contingency loci in *Neisseria meningitidis* strain MC58: evidence for a novel mechanism of phase variation. *Mol. Microbiol.* 50, 245–257
23 Falush, D. *et al.* (2001) Recombination and mutation during long-term gastric colonization by *Helicobacter pylori*: estimates of clock rates, recombination size, and minimal age. *Proc. Natl. Acad. Sci. U. S. A.* 98, 15056–15061
24 Rhen, M. *et al.* (2003) The basis of persistent bacterial infections. *Trends Microbiol.* 11, 80–86
25 Slechta, E.S. *et al.* (2003) Adaptive mutation: general mutagenesis is not a programmed response to stress but results from rare coamplification of *dinB* with *lac*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 12847–12852

0966-842X/$ - see front matter
doi:10.1016/j.tim.2003.11.008

Genome Analysis

Transporter's evolution and carbohydrate metabolic clusters

Titia H. Plantinga[1], Chris van der Does[1,2] and Arnold J.M. Driessen[1]

[1]Department of Microbiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, P.O. Box 14, 9750 AA Haren, The Netherlands
[2]Institute of Biochemistry, Biocenter, Goethe-Universität Frankfurt, Marie Curie Straße 9, D-60439 Frankfurt, Germany

The *yiaQRS* genes of *Escherichia coli* K-12 are involved in carbohydrate metabolism. Clustering of homologous genes was found throughout several unrelated bacteria. Strikingly, all four bacterial transport protein classes were found, conserving transport function but not mechanism. It appears that during evolution the ability to transport, phosphorylate and metabolize substrates of unknown identity have been conserved. However, the transporter classes have been swapped. This probably demonstrates the subtlety of transport-protein evolution.

The gene cluster designated *yiaJKLMNOPQRS* (GenBank accession number g1789999–1790008) in the *Escherichia coli* K-12 genome (Figure 1a) has been suggested to be involved in the uptake and metabolism of carbohydrates [1–5]. Recently, we have demonstrated that the binding protein-dependent secondary transporter (Figure 2a) encoded by the *yiaMNO* genes transports the rare pentose L-xylulose, providing the first evidence that these systems are able to transport carbohydrates [6]. During characterization of this transporter we became interested in the function of the remaining genes in the cluster. An increase in the availability of genome databases has permitted a search for similar gene clusters among other organisms. Homologs of the individual *yiaQ*, *yiaR* and *yiaS* genes were discovered in several bacterial and archaeal genomes (T.H. Plantinga *et al.*, unpublished). Remarkably, clustering of these genes in putative operons occurs in a limited

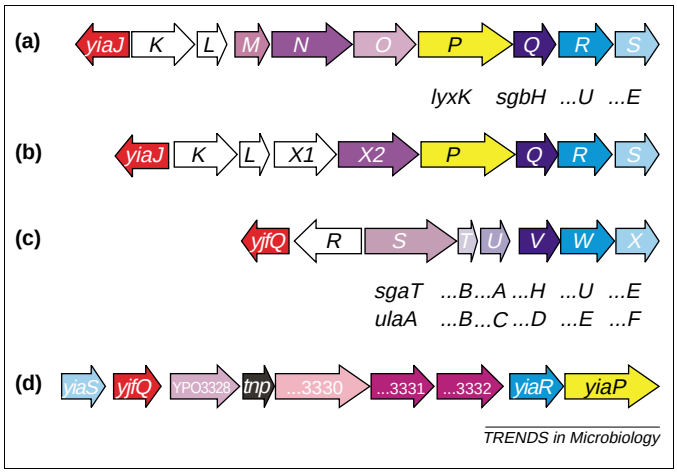

Figure 1. The structural organization of highly similar gene clusters that contain *yiaQRS* homologs (blue). These cluster with members from any of the four classes of bacterial transport proteins (purple). This is demonstrated in four model organisms. Conventional nomenclature is used, with proposed alternatives provided below each gene. Shades of blue identify genes that encode homologous enzymes: *yiaQ* and *yjfV*, 3-keto-L-gulonate 6-phosphate decarboxylases; *yiaR* and *yjfW*, (putative) L-xylulose 5-phosphate 3-epimerases; *yiaS* and *yjfX*, ribulose-5-phosphate 4-epimerases. The (putative) kinases are shown in yellow. **(a)** The *yiaJ–S* gene cluster of *Escherichia coli* K-12 contains a binding protein-dependent secondary transporter {*yiaM*, small [4 transmembrane domains (TMDs)] membrane protein; *yiaN*, large [12 TMDs] membrane protein; *yiaO*, periplasmic substrate-binding protein}. **(b)** The *yiaJ–S* gene cluster of *Klebsiella oxytoca* contains a secondary transporter (*yiaX2*) and a chemotaxis-like protein (*yiaX1*). **(c)** The *yjfQ–X* gene cluster of *E. coli* K-12 contains a phosphotransferase system (PTS) (*yjfS*, permease IIC component; *yjfT* and *yjfU*, cytosolic phosphotransferase components IIB and IIA, respectively). **(d)** The YPO3328-3332 genes of *Yersinia pestis* encode a binding protein-dependent primary ATP-binding cassette (ABC) transporter [YPO3328, periplasmic substrate-binding protein; YPO3330, ABC (ATPase) domain; YPO3331 and YPO3332, permease domains]. The gene order is not conserved, but *yiaR*, *yiaS* and *yiaP* homologs are present. Red (*yiaJ* and *yjfQ*), putative regulators; *tnp*, transposase. [1–4,7,9,17].

Corresponding author: Arnold J.M. Driessen (a.j.m.driessen@biol.rug.nl).

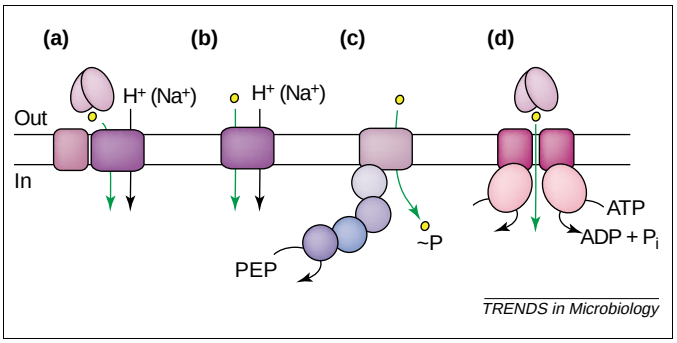

Figure 2. Domain structure of the four major classes of bacterial uptake systems. **(a)** Binding protein-dependent secondary transporters [11–14]. The substrate-binding protein delivers the substrate to the membrane-localized transporter, and unidirectional transport is driven by the proton-motive force (pmf) and/or sodium-motive force (smf). The carrier domains are dissimilar: the small domain contains four putative transmembrane domains (TMDs) and shows no homology with known membrane proteins, whereas the large domain resembles classical secondary transporters; it contains 12 putative TMDs and a large central cytosolic loop. **(b)** Secondary transporters facilitate solute uptake in a pmf- and/or smf-dependent manner [15]. A single integral membrane protein mediates reversible transport. **(c)** Phosphoenolpyruvate:carbohydrate phosphotransferase systems (PTS) use multiple cytosolic components, such as the IIA and IIB proteins, to transfer the phosphoryl group from phosphoenolpyruvate (PEP) to the carbohydrate substrate that is taken up via the IIC membrane domain [10]. **(d)** Binding protein-dependent primary transporters are members of the ATP-binding cassette (ABC) transporter superfamily [16]. The binding protein donates the solute to two identical or homologous membrane-localized transport domains. Subsequently, transport is driven by hydrolysis of ATP, which is mediated by two identical or homologous ATPases that are associated with the carrier at the cytoplasmic side of the membrane.

number of bacterial genomes. The conserved genomic organization suggests that these genes specify similar functions.

Clustering of putative carbohydrate-metabolizing enzymes

An extensive search of accessible genome databases has revealed 30 gene clusters containing putative enzymes that are highly similar to those encoded by the *yiaQRS* genes, and in most cases the gene order is well conserved (Figure 1). Interestingly, representatives are found in several unrelated Gram-positive and -negative bacterial species, most of which are known human pathogens (Table 1). Clustering of these three enzymes is found in 26 completely sequenced bacterial genomes that are available at this time, and also in the relatively small genomes of three *Mycoplasma* species (Table 1). Although homologs of the individual genes are found in the Archaea, the cluster does not appear to be conserved in members of this kingdom.

To date, only the *E. coli* K-12 enzymes have been biochemically characterized in detail. YiaQ has been suggested to function as a 3-keto-L-gulonate 6-phosphate decarboxylase based on a comparison with its homolog YjfV [7]. Because of amino acid sequence homology, YiaQ has also been suggested to function as a hexulose 6-phosphate synthase [5], but this activity has not been demonstrated experimentally [8]. The YiaR amino acid sequence shows a high similarity to 3-epimerases, and

Table 1. Bacterial genomes containing the putative carbohydrate metabolism cluster

Organism and strain	Locus	Transporter			
		BPD sec[a]	Secondary[b]	BPD ABC[c]	PTS[d]
Gram-negative:					
Escherichia coli K-12 (MG1655)	AE000435	+	−	−	+
E. coli O157:H7 (EDL933 and Sakai)	AE005174/BA000007	−	−	−	+
E. coli CFT073	AE016768	+	−	−	+
Salmonella typhimurium LT2	AE008870	+	−	−	+
S. enterica subsp. enterica serovar Typhi (CT18)	AL513382	+[e]	−	−	−
Klebsiella oxytoca	AF282849	−	+	−	−
Shigella flexneri (2a str. 301)	AE015441	−	−	−	+
Yersinia pestis (CO92)	AL590842	−	−	+[f]	−
Haemophilus influenzae Rd	L42023	+	−	−	−
H. somnus (129 PT)	NZ_AABO02000002	+	−	−	−
Pasteurella multocida	AE004439	+[f]	−	−	+[g]
Vibrio cholerae El Tor	AE003853	−	−	−	+[h]
V. vulnificus (CMCP6)	AE016811	−	−	−	+
Gram-positive:					
Mycoplasma penetrans	AP004173	−	−	−	+[i]
M. pneumoniae (M129)	U00089	−	−	−	+[i]
M. pulmonis	AL445565	−	−	−	+[i]
Streptococcus agalactiae (2603V/R)	AE014274	−	−	−	+[i]
S. mutans	AF397165	−	−	−	+[i]
S. pneumoniae (R6 and TIGR4)	AE007317/AE005672	−	−	−	+[i]
S. pyogenes (M1 GAS and MGAS8232)	AE004092/AE009949	−	−	−	+[i]
Enterococcus faecium	NZ_AAAK01000214	−	−	−	+[i]
Leuconostoc mesenteroides subsp. mesenteroides	NZ_AAABH02000028	−	−	−	+
Oceanobacillus iheyensis (HTE831)	AP004602	+	−	−	−

[a]binding protein-dependent secondary transporter.
[b]secondary transporter.
[c]binding protein-dependent ATP-binding cassette (ABC) transporter.
[d]phosphotransferase system (PTS).
[e]*yiaO* interrupted by IS-element.
[f]mixed gene order.
[g]no *yjfVWX* homologs.
[h]no IIB component, mixed gene order.
[i]no *yjfQR*.

has therefore been tentatively proposed to function as a L-xylulose 5-phosphate 3-epimerase [4]. YiaS is a ribulose 5-phosphate 4-epimerase [4,7].

The *yjfVWX* genes of *E. coli* K-12 (Figure 1c) are highly similar in sequence to the *yiaQRS* genes [5,7]. The YjfV protein is 46% identical to YiaQ and, as mentioned previously, has been shown to function as a 3-keto-L-gulonate 6-phosphate decarboxylase [7]. YjfW is a L-xylulose 5-phosphate 3-epimerase, and is 56% identical to YiaR [7]. Finally, YjfX is a ribulose 5-phosphate 4-epimerase, and 61% identical to YiaS [7]. The *yjfQ–X* cluster of *E. coli* K-12 (Figure 1c) has recently been shown to facilitate the uptake and subsequent metabolism of L-ascorbate [7,9].

Further detailed functional information concerning the gene clusters discovered in our search is lacking at this time, and the grounds for their strong conservation in a range of potential human pathogens is unknown. This intriguing question awaits further investigation.

Conservation of transport function

The potential roles that these enzyme clusters might have in carbohydrate metabolism have been described [5], but a striking feature of these systems has been overlooked up to now. It appears that the clustering of the genes encoding these enzymes coincides with the presence of a transport system (Figure 1), the class of which varies (Table 1; Figure 1). A representative from each of the four classes of prokaryotic solute-uptake systems that are known to date (Figure 2) can be found, and corresponding gene clusters are described here.

The most widespread cluster contains a PTS (phosphoenolpyruvate:carbohydrate phosphotransferase system) [5,10] (Figure 1c: '*yjfSTU*'; Figure 2c) and is found in 20 of the 30 gene clusters discovered in our search. This cluster is also the predominant one found in Gram-positive bacteria (Table 1). These systems use multiple cytosolic proteins, such as the IIA and IIB components, to transfer the phosphoryl group from phosphoenolpyruvate (PEP) to the substrate that is taken up via the permease IIC component (Figure 1c and 2c) [10]. These systems are only found in bacteria and mainly take up carbohydrates [10], although the YjfSTU system of *E. coli* was shown to transport L-ascorbate [5].

The *yiaJ–S* cluster of *E. coli* K-12 contains a binding protein-dependent secondary transporter [11,12], or a tripartite ATP-independent periplasmic (TRAP) transporter [13,14] (Figure 1a: '*yiaMNO*'; Figure 2a). In these types of transporters, the unidirectional uptake of solutes depends on the solute-binding protein, which delivers the substrate from the extracellular side to the two membrane-localized permease domains (Figure 2a). Transport is driven by the proton- or sodium-motive force (pmf or smf, respectively). The membrane proteins are dissimilar; the small permease component contains four putative transmembrane domains (TMDs) and shows no homology with known transport proteins. The large domain resembles the classical secondary transporters, containing 12 putative TMDs and a large central cytosolic loop that separates TMD 6 from TMD 7 [11–14]. Remarkably, an almost identical cluster was found in *Klebsiella oxytoca*, but the three genes encoding the YiaMNO transporter appear to have been swapped for two different genes: one encoding a secondary transporter (Figure 1b: '*yiaX2*'; Figure 2b), and one gene of unknown function that is homologous to a chemotaxis protein (Figure 1b: '*yiaX1*'). Secondary transporters facilitate solute uptake in a pmf- or smf-dependent manner [15]. Transport is reversible and is mediated by a single integral membrane protein that generally consists of 12 TMDs, with a large cytoplasmic loop between TMD 6 and TMD 7 [15] (Figure 2b).

Finally, a binding protein-dependent ATP-binding cassette (ABC) transporter (Figure 2d) has been found [16]. In *Yersinia pestis*, a *yiaR* and a *yiaS* homolog have been found in a gene cluster with such a transport system (Figure 1d: 'YPO3328–3332'). In these transporters, the solute-binding protein donates the solute to two identical or homologous membrane-localized transport domains. Uptake is driven by hydrolysis of ATP, and is mediated by two identical or homologous ATPases that are associated with the carrier on the cytosolic side of the membrane [16] (Figure 2d). The *Yersinia* cluster also contains a kinase that is homologous to *yiaP*, and a DeoR-like regulator (Figure 1d: '*yjfQ*') that is homologous to the one located upstream of the YjfSTU PTS in *E. coli* K-12 [9,17]. Although the gene order at this locus is not conserved, this localization suggests that these genes might have similar functions.

Conservation of phosphorylation function

Notably, the capacity of these gene clusters to phosphorylate the transported substrate has been conserved in all cases. The widespread PTS phosphorylates its carbohydrate substrate during transport [10] (Figure 2c). A kinase that is able to phosphorylate the substrate is located next to the three other types of transporters (Figure 1). For example, the YiaMNO transporter of *E. coli* K-12 transports L-xylulose [6], whereas, the *yiaP* gene encodes a L-xylulose kinase [1]. It is therefore probable that these systems are involved in the uptake and metabolism of a carbohydrate.

Conclusion

Taken together, 26 of those bacterial genomes that are available contain a highly similar *yiaQRS* gene cluster that probably specifies functions in carbohydrate metabolism. The gene order and the ability to phosphorylate the substrate is remarkably conserved, indicating that strong selection pressures are keeping these gene clusters together [18]. This suggests that the gene products play a crucial role in survival of the cell, and therefore might be involved in cellular processes other than carbon-source uptake and metabolism [18]. Strikingly, these clusters always contain a member from one of the four major classes of prokaryotic solute-uptake systems, which appear to have been swapped between gene clusters. It is known that transporters with different mechanisms of action can feed into the same metabolic pathways, and that many sugars are taken up by more than one transporter of different types. This has been interpreted as being a reflection of the physiological conditions. Here, we show for the first

time a case in which different transport mechanisms are genetically linked to a conserved carbohydrate metabolic cluster. Even transporters of the unusual class of binding protein-dependent secondary transporters were found linked to this cluster. It was suggested that the different classes of transporters have evolved from secondary transporters by association with binding proteins, ATPases and phosphorylating enzymes, thereby gaining higher substrate affinities and translocation power [12]. It is possible that in this cluster we can see the subtlety by which evolution can take place at the genome level.

Acknowledgements

We thank Wil Konings for helpful discussion. T.H.P. was supported by the Netherlands Organization for Scientific Research (NWO) grant no. 805–19–046 P.

References

1 Sánchez, J.C. *et al.* (1994) Activation of a cryptic gene encoding a kinase for L-xylulose opens a new pathway for the utilization of L-lyxose by *Escherichia coli*. *J. Biol. Chem.* 269, 29665–29669
2 Badía, J. *et al.* (1998) A rare 920-kilobase chromosomal inversion mediated by IS1 transposition causes constitutive expression of the *yiaK–S* operon for carbohydrate utilization in *Escherichia coli*. *J. Biol. Chem.* 273, 8376–8381
3 Ibáñez, E. *et al.* (2000) Regulation of expression of the *yiaKLMNOPQRS* operon for carbohydrate utilization in *Escherichia coli*: involvement of the main transcriptional factors. *J. Bacteriol.* 182, 4617–4624
4 Ibáñez, E. *et al.* (2000) Role of the *yiaR* and *yiaS* genes of *Escherichia coli* in metabolism of endogenously formed L-xylulose. *J. Bacteriol.* 182, 4625–4627
5 Reizer, J. *et al.* (1997) Is the ribulose monophosphate pathway widely distributed in bacteria? *Microbiology* 143, 2519–2520
6 Plantinga, T.H. *et al.* (2003) Functional characterization of the *Escherichia coli* K-12 *yiaMNO* transport protein genes. *Mol. Mem. Biol.* 10.1080/09687680310001607369 (www.tandf.co.uk)
7 Yew, W.S. and Gerlt, J.A. (2002) Utilization of L-ascorbate by *Escherichia coli* K-12: assignments of functions to products of the *yjf-sga* and *yia-sgb* operons. *J. Bacteriol.* 184, 302–306
8 Yasueda, H. *et al.* (1999) *Bacillus subtillis yckG* and *yckF* encode two key enzymes of the ribulose monophosphate pathway used by methylotrophs, and *yckH* is required for their expression. *J. Bacteriol.* 181, 7154–7160
9 Zhang, Z. *et al.* (2003) The ascorbate transporter of *Escherichia coli*. *J. Bacteriol.* 185, 2243–2250
10 Postma, P.W. *et al.* (1996) Phosphoenolpyruvate:carbohydrate phosphotransferase systems. In *Escherichia coli and Salmonella: Cellular and Molecular Biology* (Neidhardt, F.C. *et al.*, eds), pp. 1149–1174, ASM Press
11 Driessen, A.J.M. *et al.* (1997) A new family of prokaryotic transport proteins: binding protein-dependent secondary transporters. *Mol. Microbiol.* 24, 879–883
12 Driessen, A.J.M. *et al.* (2000) Diversity of transport mechanisms: common structural principles. *Trends Biochem. Sci.* 25, 397–401
13 Rabus, R. *et al.* (1999) TRAP transporters: an ancient family of extracytoplasmic solute-receptor-dependent secondary active transporters. *Microbiology* 145, 3431–3445
14 Kelly, D.J. and Thomas, G.H. (2001) The tri-partite ATP-independent periplasmic (TRAP) transporters of bacteria and archaea. *FEMS Microbiol. Rev.* 25, 405–424
15 Poolman, B. and Konings, W.N. (1993) Secondary solute transport in bacteria. *Biochim. Biophys. Acta* 1183, 5–39
16 Higgins, C.F. (2001) ABC transporters: physiology, structure and mechanism - an overview. *Res. Microbiol.* 152, 205–210
17 Campos, E. *et al.* (2002) The gene *yjfQ* encodes the repressor of the *yjfR-X* regulon (*ula*) which is involved in L-ascorbate metabolism in *Escherichia coli*. *J. Bacteriol.* 184, 6065–6068
18 Tamames, J. (2001) Evolution of gene order conservation in prokaryotes. *Genome Biol.* RESEARCH0020 (www.genomebiology.com)

0966-842X/$ - see front matter
doi:10.1016/j.tim.2003.11.001

Microbial Genomics

All things great and small

Claire M. Fraser

The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, MD 20850, USA

The report in 2002 of a novel archaeal species, the first representative of an unknown phylum, reminded us of the tremendous microbial diversity that remains to be discovered on this planet [1]. *Nanoarchaeum equitans* (the tiny archaea that rides the fire ball) is a small spherical cell that grows attached to the surface of its archaeal host, *Igniococcus* spp. These organisms were isolated from a hot submarine vent in the Kolbeinsey ridge, north of Iceland, and can be grown in the laboratory in anaerobic cultures at 90 °C in the presence of S, H_2 and CO_2. It was by microscopic examination of the *Igniococcus* cells that their 'rider', *N. equitans*, was discovered and by which it was subsequently demonstrated that *N. equitans* is an obligate symbiont of *Igniococcus*spp. that cannot be grown alone in culture. Equally surprising was that 'universal' primers designed to amplify ssRNA genes from bacterial and archaeal species failed to yield a product from *N. equitans* by the polymerase chain reaction (PCR). Subsequent comparison of the ssRNA gene sequence from *N. equitans* with those of other archeal species identified it as the first member of a novel phylum, the Nanoarchaeota, which represents the most deeply branching archaeal lineage to date. Estimation of genome size by pulse field gel electrophoresis suggested that *N. equitans* represents a minimal organism with a genome of ~500 kb.

The much anticipated genomic sequence of *N. equitans* was published recently [2]. Its genome size of 490 885 bp makes this the smallest prokaryotic genome deciphered to date. Despite its small genome size, this archaeon has one of the highest gene densities reported to date, with ~95% of the genome representing predicted coding sequences and stable RNAs. Most of the genes involved in information processing (DNA replication, transcription and translation) show most similarity to their counterparts in

Corresponding author: Claire M. Fraser (cmfraser@tigr.org).